Statistical OT Lecture 5: Duality

1 Some Motivating Questions

Brenier's theorem tells us that under some conditions any optimal coupling γ_0 is induced by the gradient of a convex function, i.e. if $X \sim \mu$, μ, ν have finite second moments, then γ_0 is the distribution of a pair $(X, \nabla\varphi_0(X))$ for some convex function φ_0.

Can we say more?

1. Is there a converse? If $\nabla\varphi$ pushes μ onto ν, and φ is convex, then is $\nabla\varphi$ an OT map? Is it unique?

2. Is there a way to directly find a "suitable" φ_0? Perhaps, by solving some nice optimization problem?

3. Given a convex function φ, is there a way to certify if $\nabla\varphi$ is an OT map between μ and ν?

Outside of the settings where Brenier's theorem applies, we still might want tools to understand the OT problem better. Duality is a powerful lens for understanding the OT program. We have already remarked that Kantorovich's formulation is an LP, and LPs always have an associated dual LP, and a property known as "strong duality" often holds. When strong duality holds it is often the case that one can gain a lot of insight into primal objects by studying the dual.

2 Discrete Case

The discrete case is once again the most transparent setting to consider. Suppose that $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$. Recall, the Kantorovich program was to find an optimal coupling γ_0:

$$\begin{aligned}
&\min_{\gamma} \sum_{i=1}^n \sum_{j=1}^n c_{ij}\gamma_{ij}, \\
\text{subject to} \quad & \gamma_{ij} \geq 0, \\
&\sum_{j=1}^n \gamma_{ij} = \frac{1}{n}, \quad \text{for } i \in \{1, \ldots, n\} \\
&\sum_{i=1}^n \gamma_{ij} = \frac{1}{n}, \quad \text{for } j \in \{1, \ldots, n\}.
\end{aligned}$$

To derive the dual LP we form the Lagrangian, and then appeal to a minimax theorem to allow us to interchange min and max. Concretely, we introduce Lagrange multipliers for each of the constraints and introduce the Lagrangian:

$$\mathcal{L}(\gamma, \mu, \alpha, \beta) := \sum_{i=1}^n \sum_{j=1}^n c_{ij}\gamma_{ij} - \sum_{i=1}^n \sum_{j=1}^n \mu_{ij}\gamma_{ij} - \sum_{i=1}^n \alpha_i \sum_{j=1}^n \left[\gamma_{ij} - \frac{1}{n}\right] - \sum_{j=1}^n \beta_j \sum_{i=1}^n \left[\gamma_{ij} - \frac{1}{n}\right].$$

Now, we can see that we could equivalently obtain the primal problem by considering:

$$\min_{\gamma} \max_{\mu \geq 0, \alpha, \beta} \mathcal{L}(\gamma, \mu, \alpha, \beta),$$

and swapping the min and max we obtain the dual problem:

$$\max_{\mu \geq 0, \alpha, \beta} \min_{\gamma} \mathcal{L}(\gamma, \mu, \alpha, \beta).$$

The minimization over γ can be explicitly carried out to obtain the problem:

$$\max_{\mu \geq 0, \alpha, \beta} \frac{1}{n} \sum_{i=1}^{n} \alpha_i + \frac{1}{n} \sum_{j=1}^{n} \beta_j,$$
$$\text{subject to } \alpha_i + \beta_j + \mu_{ij} = c_{ij},$$

and eliminating the μ variables we obtain the dual program:

$$\max_{\alpha, \beta} \frac{1}{n} \sum_{i=1}^{n} \alpha_i + \frac{1}{n} \sum_{j=1}^{n} \beta_j,$$
$$\text{subject to } \alpha_i + \beta_j \leq c_{ij},$$

An immediate observation is that the dual program always lower bounds the primal program in value, i.e.:

$$\min_{\gamma} \max_{\mu \geq 0, \alpha, \beta} \mathcal{L}(\gamma, \mu, \alpha, \beta) \geq \max_{\mu \geq 0, \alpha, \beta} \min_{\gamma} \mathcal{L}(\gamma, \mu, \alpha, \beta).$$

This fact is known as "weak duality" and always holds. A deep (and possibly surprising) fact is that these values are in fact equal in this setting (and also far more generally...) – this fact is called "strong duality". There are many ways to verify that strong duality holds in this case (appeal to von Neumann's minimax theorem, or check Slater's conditions which would tell us that for a finite-dimensional LP strong duality holds if either the primal or dual is feasible).

Interpretation: It was Kantorovich who first observed that the dual problem could be given an (economic) interpretation. He gave a slightly different interpretation (look up shadow prices). This version comes from Villani/Cafferelli – the dual problem is sometimes called the shippers problem. Recall, that we wanted to transport mass from μ to ν, suppose that a shipper came to us and promised us they would handle the shipping and only charge us loading and unloading costs. To transport from location X_i the loading cost is α_i, and to transport to location Y_j the cost is β_j.

Now, for us to accept the shipper's proposition, we just need to check that the price for unloading and loading is less than the cost of transporting mass ourselves (this is encoded in the constraints). So we always accept the shipper's deal, and weak duality tells us that the cost we would have incurred to transport mass ourselves is always larger than if we handed things over to the shipper.

Strong duality is then the statement that a clever shipper can solve the dual program, and set prices in such a way so as to force us to pay him as much as we would have if we transported the mass ourselves.

3 Duality for the Squared Euclidean Cost

We can derive the Kantorovich dual in general following the same steps as above. Throughout the remainder of the lecture we will focus again on the squared Euclidean cost. In the general case, the Lagrange multipliers are functions (which live in dual space to the space of measures, i.e. they are bounded continuous functions). Let us denote by C_b the space of bounded continuous functions on $\mathbb{R}^d$. Then we may verify the following:

$$\sup_{f,g \in C_b} \left[\int f d\mu + \int g d\nu - \int (f+g) d\gamma \right] = \begin{cases} 0, & \text{if } \gamma \in \Gamma_{\mu,\nu} \\ \infty & \text{otherwise.} \end{cases}$$

Then we can re-write the primal problem as:

$$\inf_{\gamma \in \mathcal{M}_+} \int \|x-y\|^2 d\gamma + \sup_{f,g \in C_b} \left[\int f d\mu + \int g d\nu - \int (f+g) d\gamma \right],$$

where $\mathcal{M}_+$ denotes the set of positive Borel measures. As before, interchanging the inf and sup, we obtain the lower bound:

$$\sup_{f,g \in C_b} \left[\int f d\mu + \int g d\nu + \left[\inf_{\gamma \in \mathcal{M}_+} \int \left[\|x-y\|^2 - f(x) - g(y) \right] d\gamma \right] \right].$$

Now, we can verify that:

$$\inf_{\gamma \in \mathcal{M}_+} \int \left[\|x-y\|^2 - f(x) - g(y) \right] d\gamma = \begin{cases} 0 & \text{if } f(x) + g(y) \leq \|x-y\|^2 \ \text{ for all } x, y \in \mathbb{R}^d \\ -\infty & \text{otherwise} \end{cases}.$$

(If the constraint is violated at any point (x, y), we can simply take $\gamma = t\delta_{x,y}$ and send $t \to \infty$). Consequently, we obtain the dual program:

$$\sup_{f,g \in C_b} \left[\int f d\mu + \int g d\nu \right]$$
$$\text{subject to } f(x) + g(y) \leq \|x-y\|^2,$$

which lower bounds the primal value (by weak duality).

Note that once we have arrived here, we can expand the collection of functions (to obtain a better lower bound) by observing that for every function $f \in L^1(\mu)$ and $g \in L^1(\nu)$, which satisfy the pointwise constraints, $f(x) + g(y) \leq \|x-y\|^2$ (μ, ν a.e. respectively) we have that for any coupling γ:

$$\int f d\mu + \int g d\nu = \int (f(x) + g(y)) d\gamma \leq \int \|x-y\|^2 d\gamma.$$

Taking the sup over $f \in L^1(\mu)$ and $g \in L^1(\nu)$ on the left, and the inf over γ on the right we obtain the dual program:

$$\sup_{f \in L^1(\mu), g \in L^1(\nu)} \left[\int f d\mu + \int g d\nu \right]$$
$$\text{subject to } f(x) + g(y) \leq \|x-y\|^2,$$

and a proof of weak duality. It is possible to prove that strong duality holds once again by appealing to a more general minimax theorem, but we will use Brenier's theorem to construct an explicit certificate of strong duality.

4 The Fundamental Theorem

In this section, we will prove a converse result to Brenier's theorem, which will have several useful implications.

Theorem 1. Suppose that μ, ν are measures with two bounded moments such that, μ has a density, and $X \sim \mu$. Then the following statements are equivalent:

1. γ_0 is an optimal coupling.

2. There is a proper, closed convex function φ_0 such that γ_0 has the same distribution as $(X, \nabla\varphi_0(X))$.

3. Strong duality holds, i.e.

$$\int \|x-y\|^2 d\gamma_0 = \sup_{\substack{f\in L^1(\mu), g\in L^1(\nu), \\ f(x)+g(y)\leq\|x-y\|^2}} \left[\int f d\mu + \int g d\nu\right],$$

and the optimal dual potentials are: $f_0(x) = \|x\|^2 - 2\varphi_0(x)$ and $g_0(y) = \|y\|^2 - 2\varphi_0^*(y)$.

Proof. We have already shown that (1) $\implies$ (2) (this is Brenier's theorem). Now, let us show that (2) $\implies$ (3). We know that,

$$\|x - \nabla\varphi_0(x)\|^2 = \|x\|^2 + \|\nabla\varphi_0(x)\|^2 - 2x^T\nabla\varphi_0(x).$$

By the equality case of Fenchel-Young we know that,

$$\varphi_0(x) + \varphi_0^*(\nabla\varphi_0(x)) = x^T\nabla\varphi_0(x).$$

Substituting and integrating we obtain that,

$$\begin{aligned}\int \|x-y\|^2 d\gamma_0 &= \int \|x - \nabla\varphi_0(x)\|^2 d\mu \\ &= \int (\|x\|^2 - 2\varphi_0(x))d\mu + \int (\|\nabla\varphi_0(x)\|^2 - 2\varphi_0^*(\nabla\varphi_0(x)))d\mu \\ &= \int (\|x\|^2 - 2\varphi_0(x))d\mu + \int (\|y\|^2 - 2\varphi_0^*(y))d\nu,\end{aligned}$$

where the final equality uses that $\nabla\varphi_0$ pushes μ onto ν. We can thus conclude that strong duality holds, for our choice of f_0 and g_0 if we can verify that these are in fact integrable functions which satisfy the constraints. Observe that,

$$\begin{aligned}f_0(x) + g_0(y) &= \|x\|^2 + \|y\|^2 - 2(\varphi_0(x) + \varphi_0^*(y)) \\ &\leq \|x\|^2 + \|y\|^2 - 2x^Ty = \|x-y\|^2\end{aligned}$$

using the Fenchel-Young inequality. Now, to verify integrability we observe that any proper, closed convex function has an affine minorant, i.e. there is some $a \in \mathbb{R}^d$, and $b \in \mathbb{R}$ such that,

$$\varphi_0(x) \geq a^Tx + b.$$

(Take a subgradient at any point in the interior of the domain, and this will define an affine minorant.) So we obtain that,

$$f_0(x) \leq \|x\|^2 - a^Tx - b,$$

and so $\int \max\{f_0, 0\}d\mu < \infty$ (using the fact that μ has finite second moments.) The same argument applies to g_0. Also since,

$$\int f_0 d\mu + \int g_0 d\nu = \int \|x-y\|^2 d\gamma_0 \geq 0,$$

so we have that $\int f_0 d\mu \geq -\int g_0 d\nu \geq -\int \max\{g_0, 0\}d\nu > -\infty$. Finally, we observe that,

$$\int |f|d\mu = 2\int \max\{f_0, 0\}d\mu - \int f_0 d\mu < \infty,$$

so we conclude that $f_0 \in L^1(\mu)$, and similar reasoning shows that $g_0 \in L^1(\nu)$.

Let us next show that (3) $\implies$ (1). By strong duality we have that for any other coupling γ:

$$\begin{aligned}\int \|x-y\|^2 d\gamma_0 &= \int f_0 d\mu + \int g_0 d\nu \\ &= \int (f_0+g_0) d\gamma \\ &\leq \int \|x-y\|^2 d\gamma,\end{aligned}$$

which shows that γ_0 is an OT coupling.

$\square$